

# Weekly Report

## 1 Done

### 1.1 套路

系统类型文章可视化及之后部分的“套路”：

#### 五、可视设计（Visual Design）

##### 1、设计原则（Design Rationales/Guidelines）/设计任务（Design Tasks）

- 一般与领域专家多次讨论，迭代产生

##### 2、模型（Method）

- 与可视设计密切相关的模型，比如聚类方法、图布局算法、异常探测方法等

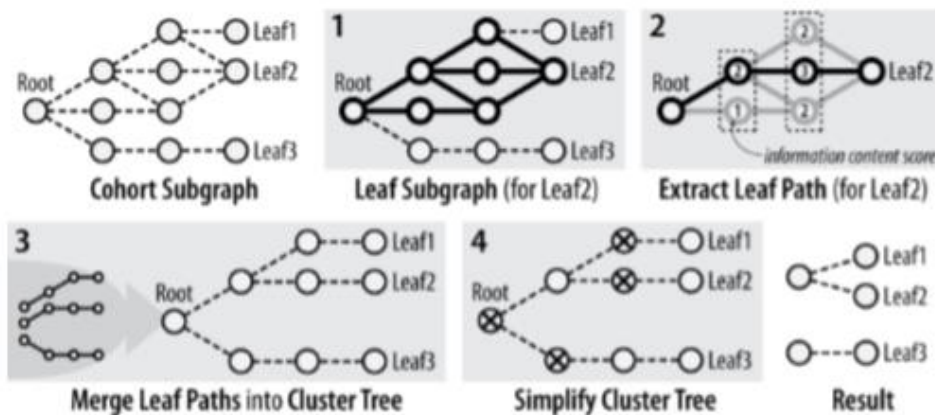


Fig. 7. Illustration of our ontology simplification algorithm.

##### 3、视图/设计（Views/Designs）

- 就视图或设计一一介绍细节。视图的介绍可以按照探索过程分组和排序。

- 总结图案（Patterns）：对应背景知识讨论可能出现的（常见的或异常的）图案及其意义

- 设计考虑：给出设计原因（可与任务/设计原则关联，显得有理有据：The UI of Voila system, as shown in Fig. 5, consists of key views corresponding to each of the tasks: (1) the macro map view shows the overview of the anomaly detection results and within the spatial context (T1); (2) the micro map view and (3) the history view respectively show the spatial and temporal context of a focal region and the associated raw information, as well as the temporal statistics from the historical data to help examine the anomaly cases (T3)...); 如有备选方案，给出讨论

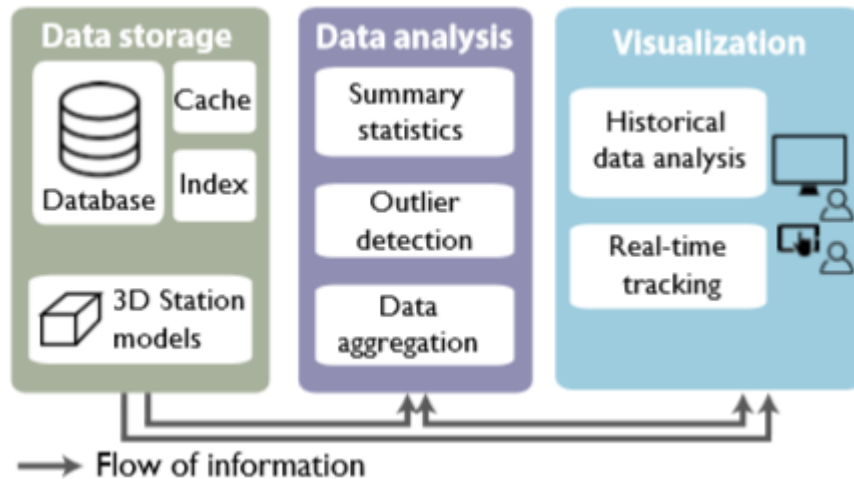
##### 4、交互（Interaction）

- 交互方式、效果。可套用经典交互分类（Navigation, filtering, linking, etc.），一一介绍它们在系统中的体现；可按照探索过程描述（**Temporal relation exploration**. To support efficient temporal relation exploration, we allow users to align sequences at a selected event. By default...）

- 设计考虑：可与任务/设计原则关联；

#### 六、实现（Implementation）

- 数据库



- 前后端（We implement a web application so the target users can access the visualizations more easily on different types of devices and platforms without any native software package installation. The front-end visualization is implemented with a combination of HTML5, CSS, JavaScript, the JavaScript Data-Driven Documents (D3) library [7], the Three.js1 WebGL library for 3D model rendering and faster...）

## 七、评估（Evaluation）

### 1、案例分析（Case study）

- 介绍数据，包括内容、数据量
- 运行环境，包括前后端、硬件配置等
- 2~3 个案例，案例描述中需要介绍操作目的、过程、发现；之前提到的功能、特点、总结的任务等需要在这里有所体现
- 用户反馈

### 2、专家采访（Expert interview）

- 专家背景，包括领域、经验（One of them focuses on designing livable public space (denoted as EA), and the other is an urban ecologist aiming at improving greenery in cities (denoted as EB). Hence, the experts are from different backgrounds: SR and EA in urban planning, while EB in ecology.）不同（相关）领域的专家可以给出多角度的评估。
- 沟通方式（面对面还是发送资料）、采访过程（介绍了多长时间，怎么介绍的，演示案例还是其他方式，专家有没有实际操作系统等）
- 专家使用背景：有什么问题，之前是怎么解决的，哪里不好
- 采访中的行为：在学习系统时遇到了什么困难（或者快速理解、掌握）；使用了什么视图，如何使用
- 观点：每个视图、交互设计的是否合理，特别是在贡献中强调过的创新；提出了什么问题；有什么建议等。**\*注意：**这部分在描述的时候一般需要先提炼出几个条目，再详细解释细节。简单地说好没有意义，需要指出好在哪里或者为什么喜欢，对已有工作有什么帮助等。是否愿意使用和是否愿意推荐给别人都是比较有说服力的评价标准。“Finally, all the researchers and research associates wanted a longitudinal version of this tool, since there were aspects of their patients that they wanted to visually track over time.” 如有必要，可以对专家的原话进行引用。

（EA highlighted “it is very important to use the same colors in different views”.）

### 3、用户采访（User Interview）：相较于 expert interview 人数更多，参与人员要求略低。可设

计先导实验以使流程更加合理。

4、定量评估（Quantitative Evaluation）/方法比较

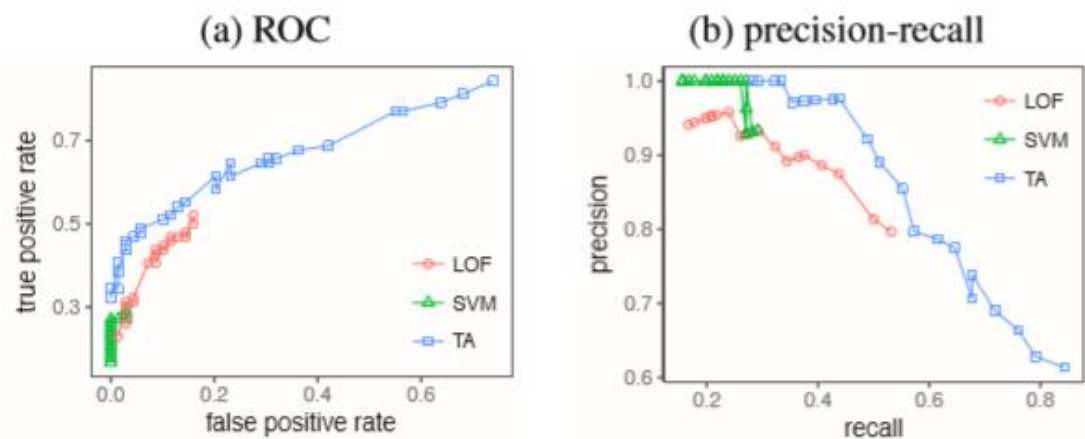


Fig. 7. Performance evaluation of anomaly detection. Results indicate our algorithm (TA) outperforms baseline methods (LOF and One-Class SVM).

- 八、讨论（Discussion）
- 优势：通过对比（和自动算法、已有方法等对比）得出本方法的优势
  - 一般性（General Application）：可套用其他场景，需要举出具体例子
  - 局限（limitation）：可以处理的数据量级（scalability）、能否覆盖用户的全部需求等，可以针对专家采访的结论讨论
  - 未来工作（Future work）：需要解决但又解决不了的问题，最好可以给出大致方向和解决思路
- 九、总结（Conclusion）
- 工作做了什么，特色是什么，解决了什么问题。

1.2 项目思路总结及进度

思路	1.抽象出“骨架”进行重构		2.分裂出“材料”进行重组
骨架/材料	二叉树（xiaoDP）	多叉树	dK-serise（salaDP）
方法	随机生成层次随机图(HRG),再基于马尔科夫链优化	社团分割生成多叉子树	按边两端节点度数，对边进行分组统计
中间结果	不确定邻接矩阵（任意两点之间有边的概率）		统计结果，矩阵形式
备注	已理清思路，正在实现。	我们的方法，推测可以更好地描述大部分图的特征。社团分割部分已实现。	已实现。

以上部分为利用差分隐私来处理图数据的方法。我们在思路 1 的基础上给出了拓展方法。思路 2 暂时没有想到可以拓展的方式。

在我们的系统中，三种方法的处理过程会通过可视化被解释处理，以辅助用户理解模型，并调节参数（参数定义在差分隐私中是一个重要问题）。除此

之外，我们将参照已有方法的评估方式，从数值和图表两方面给出评价。

### 1.3 软文-还在担心隐私泄露？让可视化来做你的保护侠！

#### 前言

不久前的 Facebook 事件引发了大众对隐私问题的关注。如何为用户的隐私构建起强大的保护再次成为热点问题。针对图数据，浙江大学，加州大学戴维斯分校和阿里云联合发表了最新研究成果图保护器（GraphProtector）—— 一个可以针对不同攻击提供有效保护的可视分析系统。

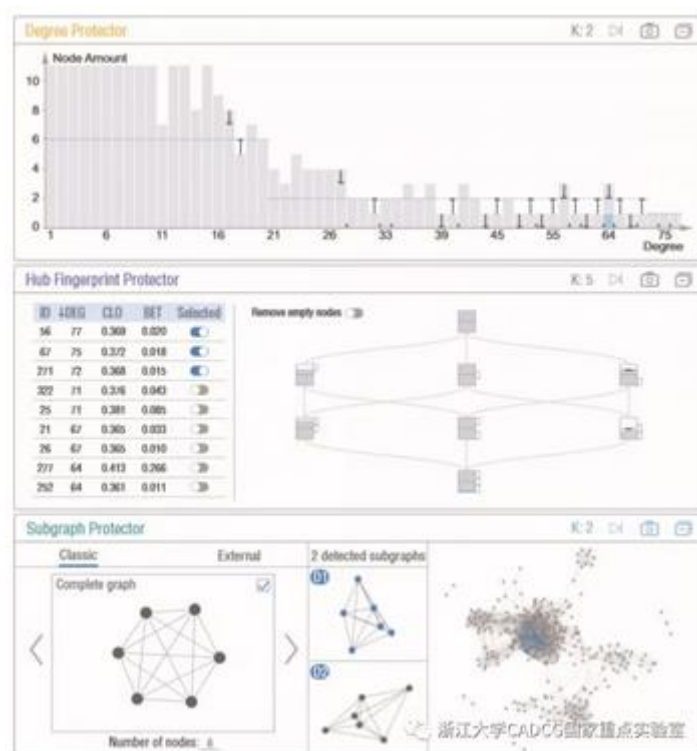
图数据是关系信息的载体，它通过节点和边来表达个体间的关系。这种数据被广泛应用于社团探测、信息传播、欺诈追踪等研究。发布涉及人与人之间关系的图数据可以推动社会学等学科的发展，但也伴随着隐私泄露的风险。利用图数据的结构特征，攻击者有机会破解用户的身份信息。在介绍图保护器前，让我们先来了解图数据的一些结构特征。

**度数：**节点的度数是和它相连的边的总数。

**关键节点指纹：**关键节点是指在图数据有明显特征的点，它们可能有很高的度数，或是连接多个社团的枢纽等。这些节点十分容易暴露，且匿名代价巨大。同时，其他节点和这些节点的连接关系也构成了一种特殊的特征，我们将它称为关键节点指纹。

**子图：**相邻的一组节点间的连接关系构成了子图，子图的规模、图案都有很多种变化。

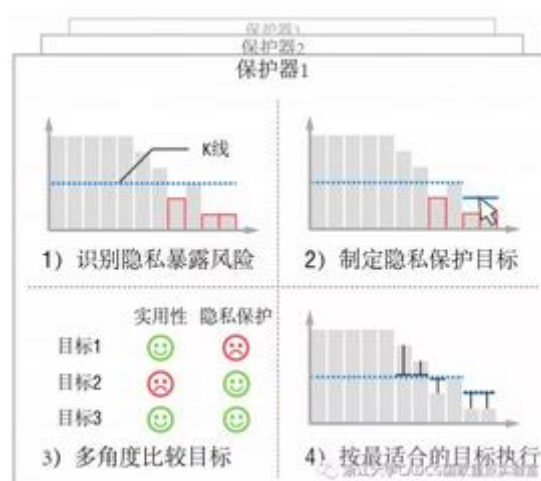
图保护器可以辅助用户基于 k-匿名方法对以上三种结构特征导致的隐私问题进行处理。K-匿名方法是一种经典的隐私保护方法。它的思路是通过构造 k 个以上相似结构特征的方式来对可能被识别出的节点进行匿名保护。



图一：针对不同结构特征的保护器

在保护的过程中， $k$  值对应了相似结构的总数，它的设定与保护级别的高低有着密切的关系。需要注意的是，需要提供的保护级别越高，处理数据时需要做出的改变就越大，随之而来的就是对数据实用性破坏的增加。因此，如何设定  $k$  是一个重要的问题。

图保护器系统允许用户基于  $k$  值提供不同的隐私保护方案。针对每个方案，辅助用户从隐私保护和实用性损失两方面进行评估，并最终得出合适的方案。



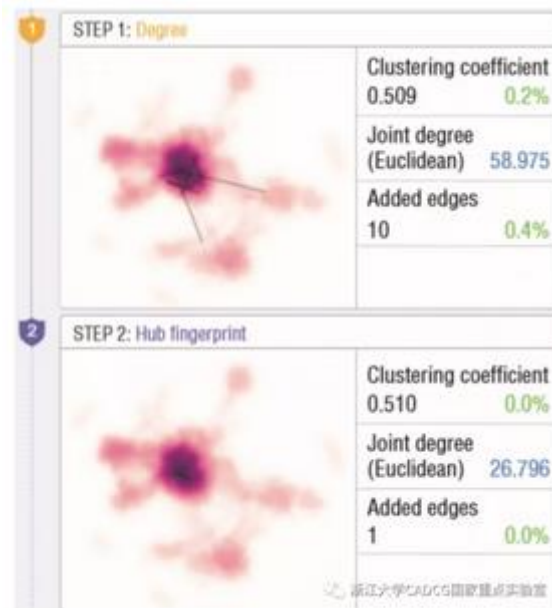
图二：保护器使用流程

在有了用户定义的方案后，算法将自动完成对数据的处理。然而，在这个过程中，经常会存在有多种处理都能实现目标的情况。举个例子，比如图中原有度数为9的点1个，度数为8的点10个。当我们要对它们实现  $k=2$  的度数匿名（也就是每个度数的节点数不少于2）时，需要给一个度数为8的节点增加一条边，使其度数为9，从而得到两个度数为9的节点，实现要求的度数匿名。但是自动算法该如何从10个度数为8的节点中做出选择呢？图保护器允许用户基于节点属性对节点分组并排序，以指导自动算法在有多个方案备选时做出合适的选择。



图三：优先级处理

除此之外，图保护器可以对用户的处理进行记录。当用户针对不同隐私问题套用了多种处理后，每一步对图数据做出的改变，引发的指标变动都会被记录下来。



图四：处理记录

## 2 Progress

Item	Deadline	Current progress	Remark
Intro of GraphProtector	8.5		
Courseware revision	9.1	Sent the current version by email.	
Go abroad	11.18	Made an appointment for US visa.	
Privacy program	10.31	Implementing existed approaches.	